



exponential decay

→ $-\frac{\partial J}{\partial \theta}$

x	y	z	...		
---	---	---	-----	--	--

batch 1 2 3 4 ...

Let us consider accelerator 0.5

i x

$x + 0.5y$

$(x + 0.5y) + 0.5z$

⋮

time we make a 'checkpoint'; we save a model

when a loss is min (similar to video game) it's a best model till now

(we will discuss later)

Nester of momentum

Check point

(1) we define a model

(2) Define hyper parameters

One way to find is to recognize elbow pattern

(3) check points → when to stop the model

(4) Stopping criterion → loss function, how many times the data has been trained etc.

⇒ Dropout is only for training but not at inference.